Nonlinear Causality Inference using a Robust ERR-Based Method in the Context of Epilepsy

Marc Greige
Univ Rennes, Inserm, LTSI - UMR 1099
F-35000 Rennes, France
marc.greige@univ-rennes.fr

Ahmad Karfoul
Univ Rennes, Inserm, LTSI - UMR 1099
F-35000 Rennes, France
ahmad.karfoul@univ-rennes.fr

Isabelle Merlet
Univ Rennes, Inserm, LTSI - UMR 1099
F-35000 Rennes, France
isabelle.merlet@univ-rennes.fr

Régine Le Bouquin Jeannès
Univ Rennes, Inserm, LTSI - UMR 1099
F-35000 Rennes, France
regine.le-bouquin-jeannes@univ-rennes.fr

***Abstract*—In this paper, a robust error reduction ratio based method is proposed for the estimation of nonlinear causality among dynamic systems. The proposed method copes with the issue of spurious solutions inherent to the original ERR-based method. These spurious solutions are removed by considering a sparse representation of the model coefficient vector. This sparse representation is recovered using the well-known alternating direction method of multipliers combined with an optimal computation of the regularization parameter at each iteration using the discrepancy principle. The robust ERR-based method is evaluated in the context of inferring brain effective connectivity in epilepsy. Results obtained on simulated and real intracerebral electroencephalographic signals confirm its efficiency.**

***Index Terms*—Error Reduction Ratio, Orthogonal Least Squares, effective connectivity, ADMM**

I. Introduction

Epilepsy is a neurological disease that affects about 1% of the population worldwide. It is the fourth most common neurological disorder and concerns people of all ages. Epilepsy is generally characterized by repetitive seizures, called critical periods, which vary in frequency and duration and are the result of excessive and abnormal activity of cortical nerve cells in the brain. Around 30% of epileptic patients are drug-resistant and, in such cases, surgery can be considered. The goal of pre-surgery is to delineate the epileptic zone (EZ) that is responsible for the initiation and/or the propagation of the epileptic seizure. The organization of the epileptogenic zone often corresponds to that of a network of neuronal ensembles distributed in potentially distant structures. Thus, identifying the EZ to be delineated requires not only the identification of the neural network involved in the epileptic seizure set up but also in analyzing causal relationships among its nodes (neuronal ensembles). Intracerebral electroencephalographic (iEEG) signals are commonly used to measure the cerebral activity before and during the seizure [1], [2]. It is an invasive technique which, contrary to scalp EEG recordings, ensures a relatively high signal to noise ratio and is free from the effect of volume conduction. Neural activities result from nonlinear processes, and interactions between neural assemblies tend generally to be nonlinear rather than linear, an assumption commonly considered in effective connectivity measures derived from Granger causality index [3]. Brain effective connectivity serves to quantify causal relations among different brain regions. The error reduction ratio (ERR)-based method [4]–[8] has already shown promising results in identifying nonlinear systems and hence nonlinear effective connectivity among their compartments. It is a dictionary based method where the signal at hand is decomposed as a linear combination of some linear and/or nonlinear terms that are chosen from a predefined dictionary of possible candidate models. Such decomposition relies on the orthogonal least squares (OLS) algorithm widely used to solve nonlinear system identification problems. The retained candidate models in the initial dictionary are chosen according to an error reduction ratio measure as the ones that significantly, up to a given threshold, contribute to the signal at hand. However, despite its efficiency, the ERR-based method suffers from the presence of spurious terms whose number is subject to the ERR-related threshold. This fact will, to a large extent, affect the quality of the system identification. To cope with this limitation, the ERR-based nonlinear system identification method is revisited in this paper so that all spurious models are removed. The modified version of the ERR-based approach relies on the assumption that only few but the most significant terms among all retained candidate models by the ERR-based method are really contributing to the observed signal. This leads to a sparse representation of the estimated model coefficient vector. This sparse representation is recovered here using the well-known Alternating Direction Method of Multipliers (ADMM) [9]. This paper is organized as follows: Section II is devoted to the novel methodology we propose to reconstruct nonlinear signals. Experimental results on simulated and real iEEG signals are presented in Section III and discussed. Finally, concluding remarks are given in Section IV.

II. METHODOLOGY

Let $\boldsymbol{y}_m \in \mathbb{R}^T$ be an intracerebral EEG signal of length T recorded on the m-th, $1 \leq m \leq M$ sensor where T is the observation period and M is the number of sensors. The m-th observed iEEG signal corresponds to the activity of the m-th brain region. As brain regions are permanently communicating to set up a specific normal/pathological activity, the activity of the m-th brain region, $\boldsymbol{y}_m$, is linked to the ones of some other brain regions. More precisely, assume that $\boldsymbol{y}_m$ is decomposed as a linear combination of a set of N_m time series, denoted by $\tilde{\boldsymbol{y}}_i^{(m)}, 1 \leq i \leq N_m$. Each of the latter time series could be figured out as a function, denoted here by $f_i^{(m)}$, encoding a linear and/or nonlinear combination of a subset of delayed versions of the acquired iEEG signals $\{\boldsymbol{y}_k^{\tau_k}\}_{\forall k \in \Omega_i^{(m)}, \forall \tau_k \in \Phi_i^{(m)}}$ where the indices of these time series and their related time lags are defined in the sets $\Omega_i^{(m)}$ and $\Phi_i^{(m)}$, respectively. This can be expressed as follows:

$$\begin{aligned} \boldsymbol{y}_m &= \sum_{i=1}^{N_m} \alpha_i^{(m)} \tilde{\boldsymbol{y}}_i^{(m)} + \boldsymbol{w}_m \\ &= \sum_{i=1}^{N_m} \alpha_i^{(m)} f_i^{(m)}(\{\boldsymbol{y}_k^{\tau_k}\}_{\forall k \in \Omega_i^{(m)}, \forall \tau_k \in \Phi_i^{(m)}}) + \boldsymbol{w}_m \end{aligned} \tag{1}$$

where $\alpha_i^{(m)}$ is the i-th decomposition coefficient and $\boldsymbol{w}_m \in \mathbb{R}^T$ is the model residual related to $\boldsymbol{y}_m$. Understanding linear/nonlinear interactions among brain regions and revealing the connection strength can be summed up to the identification of not only the set of signals $\{\boldsymbol{y}_k^{\tau_k}\}_{\forall k \in \Omega_i^{(m)}, \forall \tau_k \in \Phi_i^{(m)}}$ but also to define the N_m functions $f_i^{(m)}$ and the parameter vector $\boldsymbol{\alpha}_m = [\alpha_1^{(m)}, \cdots, \alpha_{N_m}^{(m)}]^\intercal$ associated to $\boldsymbol{y}_m$. Following [5], the problem is reduced to a dictionary based decomposition of $\boldsymbol{y}_m$ as follows:

$$\boldsymbol{y}_m = \boldsymbol{D}_m \boldsymbol{\alpha}_m + \boldsymbol{w}_m, \quad \forall m \in \{1, \cdots, M\} \tag{2}$$

where $\boldsymbol{D}_m$ is a dictionary collecting the N_m times series constituting the signal $\boldsymbol{y}_m$. These times series stand for the most relevant candidates that can be selected from a known initial dictionary $\boldsymbol{D} = [\boldsymbol{d}_1, \cdots, \boldsymbol{d}_N] \in \mathbb{R}^{T \times N}$ with N being the total number of candidates. The predefined $\boldsymbol{D}$ dictionary encodes both the predefined linear and/or nonlinear functions in the set $\{\boldsymbol{y}_k^{\tau_k}\}_{\forall k \in \Omega_i^{(m)}, \forall \tau_k \in \Phi_i^{(m)}}$. More precisely, the n-th column $\boldsymbol{d}_n$ in $\boldsymbol{D}$ stands for a possible candidate of $\tilde{\boldsymbol{y}}_i, 1 \leq i \leq N_m$ in equation (1). In fact, $\boldsymbol{d}_n$ is defined as a linear and/or nonlinear mapping of a subset of predefined delayed versions of the observed iEEG signals. Then, decomposing the signal $\boldsymbol{y}_m$ is reduced to estimate the most relevant columns in $\boldsymbol{D}$, and their corresponding coefficients, that are required to fit $\boldsymbol{y}_m$ properly up to an ERR criterion [5]. The OLS algorithm is used to solve this system identification problem [4]–[6]. To this end, the matrix $\boldsymbol{D}$ is decomposed as $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{W}$ where $\boldsymbol{U} \in \mathbb{R}^{T \times N}$ and $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ are orthogonal and upper triangular matrices, respectively. For the sake of readability, the subscript m will be dropped from now on keeping in mind that the m-th, $m \in \{1, \cdots, M\}$, signal $\boldsymbol{y}_m$ is being processed. Now, in the initial stage of the ERR algorithm [5], all columns of the initial dictionary $\boldsymbol{D}$ are considered to reconstruct $\boldsymbol{y}$. Then the number of these columns is progressively reduced from N to N_m according to an ERR criterion. More precisely, let's start with:

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\theta} = \boldsymbol{U}\tilde{\boldsymbol{\theta}} = \sum_{n=1}^{N} \tilde{\theta}_n \boldsymbol{u}_n \tag{3}$$

where $\boldsymbol{\theta} \in \mathbb{R}^N$ is a coefficient vector, $\tilde{\boldsymbol{\theta}} = \boldsymbol{W}\boldsymbol{\theta}$, $\boldsymbol{u}_n$ is the n-th column of $\boldsymbol{U}$ and $\tilde{\theta}_n$ stands for the n-th component of the coefficients vector $\tilde{\boldsymbol{\theta}}$. Then decomposing $\boldsymbol{y}$ requires the identification of a subset $\Gamma = \{\boldsymbol{u}_{k_\ell}\}_{k_\ell \in \{1 \cdots, N\}, \ell \in \{1, \cdots, N_m\}}$ of the most N_m relevant column vectors of $\boldsymbol{U}$ contributing to $\boldsymbol{y}$ and their corresponding coefficients $\tilde{\theta}_\ell, 1 \leq \ell \leq N_m$. According to [4]–[6], the elements of Γ are found sequentially according to their contribution (from the highest to the lowest) to $\boldsymbol{y}$. To this end, for the sake of convenience, let $\boldsymbol{D}^{-(0)} = \boldsymbol{D}$ be the initial dictionary matrix that is used to estimate the first relevant vector in Γ, $\boldsymbol{u}_{k_1}$. Then, the matrix $\boldsymbol{D}^{-(k_i-1)} \in \mathbb{R}^{T \times N-k_i+1}$ is a reduced dictionary matrix that will be used to estimate $\boldsymbol{u}_{k_i}, k_i > 1$. The matrix $\boldsymbol{D}^{-(k_i-1)}$ is obtained by excluding from $\boldsymbol{D}^{-(k_i-2)}$ the column vector corresponding to the most relevant candidate model defining $\boldsymbol{u}_{k_i-1}$. To find this most relevant column vector in $\boldsymbol{D}^{-(k_i-1)}$, a grid search over the columns of $\boldsymbol{D}^{-(k_i-1)}$ is made. More precisely, let $\tilde{\boldsymbol{U}}_{k_i} = [\boldsymbol{u}_{k_i}^1, \cdots, \boldsymbol{u}_{k_i}^{N-k_i+1}] \in \mathbb{R}^{T \times N-k_i+1}$ be the matrix encoding the most relevant column vectors of $\boldsymbol{U}$ and defined as:

$$\tilde{\boldsymbol{U}}_{k_i} = \boldsymbol{D}^{-(k_i-1)} - \tilde{\boldsymbol{U}}_{k_i-1}\boldsymbol{H}_{k_i} \tag{4}$$

where $\tilde{\boldsymbol{U}}_{k_i-1} = \boldsymbol{u}_{k_i-1}\boldsymbol{1}^\intercal{}_{N-k_i+1}$ and $\boldsymbol{H} \in \mathbb{R}^{N-k_i+1 \times N-k_i+1}$ is a diagonal matrix that can be obtained by solving the following optimization problem:

$$\boldsymbol{H}_{k_i} = \arg\min_{\boldsymbol{H}_{k_i}} ||\boldsymbol{D}^{-(k_i-1)} - \tilde{\boldsymbol{U}}_{k_i-1}\boldsymbol{H}_{k_i}||_F^2 \; s.t. H_{k_i,\, \substack{i,j \\ \forall i \neq j}} = 0 \tag{5}$$

where $H_{k_i,i,j}$ is the (i,j)-th entry of $\boldsymbol{H}_{k_i}$ and $\boldsymbol{1}_N$ is a N-dimensional column vector of ones. Once the vector $\boldsymbol{u}_{k_i}$ is estimated, the vector $\tilde{\boldsymbol{\theta}}_{k_i}$ is computed in a least squares sense:

$$\tilde{\boldsymbol{\theta}}_{k_i} = \arg\min_{\tilde{\boldsymbol{\theta}}_{k_i}} ||\boldsymbol{y} - \tilde{\boldsymbol{U}}_{k_i}\tilde{\boldsymbol{\theta}}_{k_i}||_2^2 \tag{6}$$

Then, the $(N-k_i+1)$-dimensional ERR vector, denoted here by $\boldsymbol{e}$, is defined by:

$$\boldsymbol{e}_{k_i} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\tilde{\boldsymbol{\theta}}_{k_i}^{\odot 2} \tag{7}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the vector $[||\boldsymbol{u}_{k_i}^1||_2^2, \cdots, ||\boldsymbol{u}_{N-k_i+1}^1||_2^2]^\intercal$ as its diagonal, $\boldsymbol{\Psi} = \frac{1}{||\boldsymbol{y}||_2^2}\boldsymbol{I}_{N-k_i+1}$, $\odot$ stands for the Hadamard product (element-wise matrix product), $\tilde{\boldsymbol{\theta}}^{\odot 2} = \tilde{\boldsymbol{\theta}} \odot \tilde{\boldsymbol{\theta}}$ and $\boldsymbol{I}_K$ is a $(K \times K)$ identity matrix. Note that the ℓ-th component, e_ℓ, $1 \leq \ell \leq N-k_i+1$, of the vector $\boldsymbol{e}_{k_i}$ quantifies the contribution power of the ℓ-th candidate model, $\boldsymbol{d}_\ell^{-(k_i-1)} \in \mathbb{R}^T$, in the current dictionary $\boldsymbol{D}^{-(k_i-1)}$. Once the $N - k_i + 1$ ERR values are computed, then the index of the highest ERR value, $e_{max}^{(k_i)}$, in the vector $\boldsymbol{e}_{k_i}$ refers to the position of the most relevant candidate in $\boldsymbol{D}^{-(k_i-1)}$. The above mentioned

steps are repeated until N_m candidate models are selected for which the inequality $1 - \sum_{i=1}^{N_m} e_{\max}^{(i)} < \epsilon$, where ϵ is a predefined threshold, becomes true. Let us now define $\bar{\boldsymbol{D}} \in \mathbb{R}^{T \times N_m}$ as the dictionary collecting the N_m retained column vectors of the initial dictionary $\boldsymbol{D} \in \mathbb{R}^{T \times N}$. Then, in order to avoid some spurious retained models in $\bar{\boldsymbol{D}}$ that might appear due to the choice of the threshold ϵ, the dictionary $\bar{\boldsymbol{D}}$ should be refined. To this end, we assume that, among all retained models, few of them are relevant. This fact can be characterized through a sparse representation of the coefficient vector $\boldsymbol{\theta}$. The optimal spurious model-free vector $\boldsymbol{\theta}$ is found by solving the following optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{\lambda}{2}||\boldsymbol{y} - \boldsymbol{x}||_2^2 + ||\boldsymbol{z}||_1 \text{ s.t. } \boldsymbol{x} = \bar{\boldsymbol{D}}\boldsymbol{\theta} \text{ and } \boldsymbol{z} = \boldsymbol{\theta} \tag{8}$$

where λ is a regularization parameter and $||.||_1$ is the $L1$-norm. Such optimization problem can be solved using the ADMM which minimizes the augmented Lagrangian function [10] associated to (8) and given by:

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{v}, \boldsymbol{g}, \lambda) = \frac{\lambda}{2}||\boldsymbol{y} - \boldsymbol{x}||_2^2 + ||\boldsymbol{z}||_1 + \frac{\rho_1}{2}||\boldsymbol{\theta} - \boldsymbol{z}||_2^2 \\ + \boldsymbol{v}^\intercal(\boldsymbol{\theta} - \boldsymbol{z}) + \frac{\rho_2}{2}||\bar{\boldsymbol{D}}\boldsymbol{\theta} - \boldsymbol{x}||_2^2 + \boldsymbol{g}^\intercal(\bar{\boldsymbol{D}}\boldsymbol{\theta} - \boldsymbol{x})\end{aligned} \tag{9}$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ are auxiliary variables, $\boldsymbol{v}$ and $\boldsymbol{g}$ stand for the Lagrange multipliers and $\rho_1, \rho_2 \in \mathbb{R}_+^*$. The update rules of $\boldsymbol{x}, \boldsymbol{\theta}$ are obtained by looking for the stationary points of $\mathcal{L}$ in $\boldsymbol{x}$ and $\boldsymbol{\theta}$, respectively. This leads to :

$$\boldsymbol{\theta} = (\rho_1 \boldsymbol{I}_N + \rho_2 \bar{\boldsymbol{D}}^\intercal \bar{\boldsymbol{D}})^{-1}(\boldsymbol{v} + \rho_1 \boldsymbol{z} + \bar{\boldsymbol{D}}^\intercal(\rho_2 \boldsymbol{x} - \boldsymbol{g})) \tag{10}$$

$$\boldsymbol{x} = \frac{\lambda \boldsymbol{y} + \boldsymbol{g} + \rho_2 \bar{\boldsymbol{D}}\boldsymbol{\theta}}{\lambda + \rho_2} \tag{11}$$

Regarding the Lagrange variables, $\boldsymbol{v}$ and $\boldsymbol{g}$, they are updated using the ascent-gradient scheme:

$$\boldsymbol{v}_{i+1} = \boldsymbol{v}_i + \rho_1(\boldsymbol{\theta} - \boldsymbol{z}), \ \boldsymbol{g}_{i+1} = \boldsymbol{g}_i + \rho_2(\bar{\boldsymbol{D}}\boldsymbol{\theta} - \boldsymbol{x}) \tag{12}$$

where i is the iteration index. As far as the dual variable $\boldsymbol{z}$ is considered, it is updated as follows:

$$\boldsymbol{z} = prox_{\phi, 1/\rho_1}(\boldsymbol{\theta} + \frac{1}{\rho_1}\boldsymbol{v}) \tag{13}$$

where $prox$ is the proximal operator handling the non-smooth part ($i.e.$ $L1$-norm) in the above augmented Lagrangian function (equation (9)). It is the shrinkage operator with soft thresholding initially proposed in [9], with $\frac{1}{\rho_1}$ as the shrinking threshold. Regarding the regularization parameter λ, it is optimally computed, at each iteration, by means of the discrepancy principle [11]. More precisely, this principle considers that the regularization parameter is laying in the set $\{\boldsymbol{x} : ||\boldsymbol{x} - \boldsymbol{y}||_2^2 \leq c\}$ where $c \in \mathbb{R}$ is a coefficient relating to the noise variance [11] and can be computed by means of the equivalent degree of freedom method [12] [13]. This leads to:

$$\lambda = \frac{||\rho_2(\boldsymbol{y} - \bar{\boldsymbol{D}}\boldsymbol{\theta}) - \boldsymbol{g}||_2}{\sqrt{c}} - \rho_2 \tag{14}$$

All variables in the optimization problem are updated alternatively in an iterative way where, at each iteration, each variable is updated while keeping the other variables fixed to their last estimate. The optimization process stops either when the relative error on the estimation of the vector $\boldsymbol{\theta}$ between two successive iterations exhibits a value that is smaller than or equal to a predefined threshold, or when a maximal number of iterations is reached.

III. Experimental Results

A. Simulated data

In this section two simulated models were used to evaluate the behavior of the proposed robust ERR-based method compared to the original ERR-based one. To do so, a linear multivariate auto-regressive (MVAR) model was first considered while a nonlinear model was used afterwards. The simulated data were 1024-point long, corresponding to four seconds of iEEG signals sampled at 256 Hz. To assess the signal reconstruction quality of the proposed approach the mean squared error (MSE) is considered in this study:

$$MSE_m = \frac{1}{K}\sum_{k=1}^{K} ||\boldsymbol{y}_m - \hat{\boldsymbol{y}}_{m,k}||_2^2 \tag{15}$$

$\forall m \in \{1, \cdots, M\}$ where $\hat{\boldsymbol{y}}_{m,k}$ is the estimate of $\boldsymbol{y}_m$ at the k-th Monte Carlo trial ($1 \leq k \leq K = 1000$) and $\hat{\boldsymbol{y}} = \bar{\boldsymbol{D}}\boldsymbol{\theta}^*$ is the estimation of $\boldsymbol{y}$.

1) Linear Model: The first model is a 5-channel linear model taken from [14], which simulates neural activity over 5 brain regions. It is expressed as follows:

$$\begin{aligned}y_1(k) &= 0.6y_1(k-1) + 0.655y_2(k-2) + w_1(k) \\ y_2(k) &= 0.5y_2(k-1) - 0.3y_2(k-2) - 0.3y_3(k-4) \\ &\quad + 0.6y_4(k-1) + w_2(k) \\ y_3(k) &= 0.8y_3(k-1) - 0.7y_3(k-2) - 0.1y_5(k-3) + w_3(k) \\ y_4(k) &= 0.5y_4(k-1) + 0.9y_3(k-2) + 0.4y_5(k-2) + w_4(k) \\ y_5(k) &= 0.7y_5(k-1) - 0.5y_5(k-2) - 0.2y_3(k-1) + w_5(k)\end{aligned} \tag{16}$$

where $\boldsymbol{w}_m \sim \mathcal{N}(0, 1), 1 \leq m \leq 5$. In this case, forty candidates were collected in an initial dictionary $\boldsymbol{D}$. This dictionary was built by concatenating (i) a set of delayed versions of $\{\boldsymbol{y}_m^{\tau_m}\}_{1 \leq m \leq 5}, \forall \tau_m \in \Phi_m$ where $\Phi_m = \{1, \cdots, 5\}$ is the set of their related time lags, and (ii) the set $\{\tilde{\boldsymbol{y}}_1^{(m)}, \cdots, \tilde{\boldsymbol{y}}_{15}^{(m)}\}$ with $\tilde{\boldsymbol{y}}_i^{(m)} = f_i^{(m)}(\boldsymbol{y}_j^1, \boldsymbol{y}_l^1) = \boldsymbol{y}_j^1 \boldsymbol{y}_l^1, 1 \leq j, l \leq 5$ and $y_j^\xi(k) = y_j(k - \xi)$.

TABLE I
MSE for the '5-channel linear model'

	ERR-based method	***Proposed method***
y_1	1.862 ± 0.065	1.886 ± 0.066
y_2	5.252 ± 0.632	$\mathbf{1.434}$ ± 0.040
y_3	2.045 ± 0.203	$\mathbf{1.086}$ ± 0.039
y_4	6.336 ± 0.739	$\mathbf{1.719}$ ± 0.051
y_5	2.929 ± 0.203	$\mathbf{0.346}$ ± 0.013

It can be noticed from Table I that the signal reconstruction error is substantially lower for the proposed robust ERR-based method compared to the original ERR-based one. In order to evaluate the statistical significance of the results in the context of nonlinear system identification, a Wilcoxon signed rank test was employed. More precisely, for each signal at hand, $\boldsymbol{y}_m, m \in \{1, \cdots, M\}$ a correlation series was computed

for both methods, the proposed method (PM) and the ERR-based one respectively, leading to $\boldsymbol{g}_m^{PM}$ and $\boldsymbol{g}_m^{ERR}$ correlation series respectively. More precisely, the k-th component $g_{m,k}^{\ell}$ of the correlation series $\boldsymbol{g}_m^{\ell}$, $\ell \in \{PM, ERR\}$, stands for the correlation coefficient between the estimated signal $\hat{\boldsymbol{y}}_m$ using the ℓ-th method and the true signal $\boldsymbol{y}_m$. The Wilcoxon signed rank test was then applied to the pair ($\boldsymbol{g}_m^{PM}$, $\boldsymbol{g}_m^{ERR}$). Box plots of the difference between the two correlation time series $\boldsymbol{e}_m = \boldsymbol{g}_m^{PM} - \boldsymbol{g}_m^{ERR}$ are shown in Figure 1 together with their corresponding p-values. According to this figure, the proposed method provides a signal reconstruction that is more consistent with the ground truth compared to the conventional ERR-based algorithm. This result is true for all simulated signals described in equation (17) excepted for $\boldsymbol{y}_1$ where a tiny difference is observed ($i.e.$, quartiles $Q_i \approx 0$ for $\boldsymbol{e}_1, i \in \{1, 2, 3\}$). The superiority of the proposed approach over the original ERR-based one is also statistically confirmed using the five computed p-values depicted in Figure 1 with **** since their respective values are all less than 0.0001.

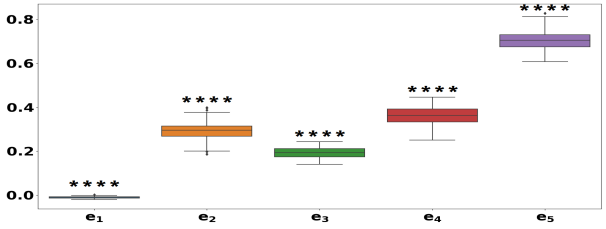

Fig. 1. Box-plots of the difference, $\boldsymbol{e}_m = \boldsymbol{g}_m^{PM} - \boldsymbol{g}_m^{ERR}, m \in \{1, \cdots, 5\}$, between paired correlation series $\boldsymbol{g}_m^{ERR}$ and $\boldsymbol{g}_m^{PM}$ for the five simulated signals described in (16). p-value ≤ 0.0001 is marked as ****.

2) Nonlinear model: The second model is nonlinear and has been proposed in [6]. This nonlinear model considers two signals. Interactions between these two signals $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ occur in the time intervals $[101; 300]$ and $[501; 700]$. In fact, a causal effect from $\boldsymbol{y}_1$ to $\boldsymbol{y}_2$ occurs in the interval $[101; 300]$ and is modeled through the following equation:

$$\begin{aligned} y_1(k) &= r(k) \\ y_2(k) &= -0.07y_1(k-1) + 0.32y_1(k-2) \\ &\quad - y_1(k-1)y_1(k-2) + w_1(k) \end{aligned} \tag{17}$$

In addition, a causal effect from $\boldsymbol{y}_2$ and $\boldsymbol{y}_1$ occurs in the interval $[501; 700]$ as follows:

$$\begin{aligned} y_1(k) &= -0.07y_2(k-1) + 0.32y_2(k-2) \\ &\quad - y_2(k-1)y_2(k-2) + w_2(k) \\ y_2(k) &= r(k) \end{aligned} \tag{18}$$

where $\boldsymbol{w}_m \sim \mathcal{N}(0, 0.1), 1 \leq m \leq 2$, $\boldsymbol{r}$ is a random data sequence that follows a uniform probability distribution over $[-1; 1]$. Regarding the other time intervals, signals $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are set equal to $\boldsymbol{r}$. Twenty-eight linear and nonlinear candidate models were initially selected as suggested in [6].

TABLE II
MSE FOR THE '2-CHANNEL NONLINEAR MODEL'

	ERR-based method	***Proposed method***
y_1	0.112 ± 0.010	$\mathbf{0.040}$ ± 0.003
y_2	0.103 ± 0.013	$\mathbf{0.090}$ ± 0.005

According to Table II, using a nonlinear model, the new algorithm outperforms the ERR-based method in terms of signal reconstruction. As previously, the Wilcoxon signed rank test was applied on the computed correlation series related to the signals described in equations (17) and (18). Results confirm again that the difference in the system identification quality, which is in favor of the proposed approach as already shown in Table II, is statistically significant for the two considered signals ($i.e.$, p-values ≤ 0.0001).

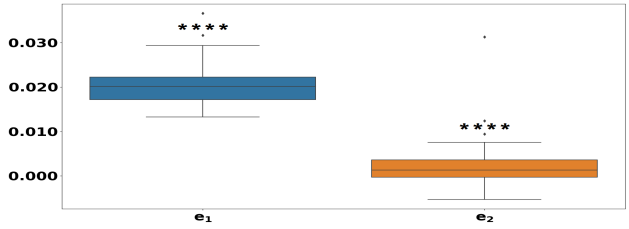

Fig. 2. Box-plots of the difference, $\boldsymbol{e}_m = \boldsymbol{g}_m^{PM} - \boldsymbol{g}_m^{ERR}, m \in \{1, 2\}$, between paired correlation series $\boldsymbol{g}_m^{ERR}$ and $\boldsymbol{g}_m^{PM}$ for the two simulated signals described in (17), (18). p-value ≤ 0.0001 is marked as ****.

B. Real iEEG data

Real iEEG data were acquired from a female patient, aged 35, at the Epilepsy Unit in Rennes Hospital, France. Twelve intracerebral electrodes (10-15 contacts) were implanted in the left temporal, insular, inferior frontal and inferior parietal regions. A 76-second length portion of recording, sampled at 256 Hz, was considered. We kept only 12 bipolar channels, corresponding to the brain regions that were the most involved in the present seizure according to preliminary clinical and electrophysiological examinations by the clinician. The expert classified the twelve channels into three different groups. First, the 'Onset' group (O) is the group containing the channels that were majorly responsible for the initiation of the seizure since rapid discharges were observed in these channels. Secondly, the 'Propagation Internal' group (P_I) contains the channels related to the regions that are triggered by the O group. This group can be marginally involved in the seizure through delayed electrical discharges with lower intensity in comparison with the channels belonging to the O group. Hence, the P_I group refers to brain regions that are less epileptogenic, and therefore considered as the one linking the most epileptogenic zones to those which are the less epileptogenic. Thirdly, the 'Propagation Sink' group (P_S) consists of the brain regions which are mainly triggered by the O group, and therefore considered as not responsible for the triggering of the seizure. According to the clinical expert, the most interesting time period to be considered corresponds to the onset of the ictal phase, which was estimated to be around the 20th second. To detect fast transition in our case, we selected the interval $[18s; 22s]$ just before the beginning of the seizure where we expected a change in the behaviour of the studied brain regions.

To classify the channels, each real iEEG signal, $\boldsymbol{y}_m$, $1 \leq m \leq M$ ($M = 12$) is assigned to either the O, P_I or P_S group using a defined threshold ϕ_{th}:

$$\phi_{th} = \frac{1}{2M} \sum_{m=1}^{M} |\phi_m| \tag{19}$$

where ϕ_m is defined by $\phi_m = \frac{OD_m - ID_m}{OD_m + ID_m}$ with OD_m and ID_m stand respectively for the outward and the inward degrees of the m-th signal (node) in the estimated brain network. More precisely, let $\mathbf{\Theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_M] \in \mathbb{R}^{M \times M}$ be the adjacency matrix associated to the directed graph describing the estimated brain network. Then, we have [15]:

$$OD_m = \sum_{i=1}^{M} \Theta_{m,i} \quad , \quad ID_m = \sum_{i=1}^{M} \Theta_{i,m} \tag{20}$$

where $\Theta_{m,i}$ denotes the (m,i)-th entry of $\mathbf{\Theta}$. It is noteworthy that the adjacency matrix associated to a directed graph is a square asymmetric matrix ($i.e.$, $\Theta_{i,j} \neq \Theta_{j,i}$). Thus, the classification rule for a given signal $\boldsymbol{y}_m$ is defined by:

$$\boldsymbol{y}_m \in \begin{cases} O, & \text{if } \phi_m \geq \phi_{th} \\ P_I, & \text{if } -\phi_{th} \leq \phi_m \leq \phi_{th} \\ P_S, & \text{if } \phi_m \leq -\phi_{th} \end{cases} \tag{21}$$

The clinician's classification is reported in Table III. The results obtained by the ERR-based method and the proposed one are given in Table IV where the correctly classified channels are written in bold.

TABLE III
EXPERT'S CLASSIFICATION

O group	P_I group	P_S group
$Bp1$-$Bp2$	$Bp6$-$Bp7$	$Fp2$-$Fp3$
$Cp1$-$Cp2$	$Ap6$-$Ap7$	$Tp1$-$Tp2$
$Pp1$-$Pp2$	$Dp1$-$Dp2$	
$Pp4$-$Pp5$	$Cp4$-$Cp5$	
$Pp8$-$Pp9$		
$Ap2$-$Ap3$		

TABLE IV
CLASSIFICATION USING (A) THE ERR-BASED METHOD, (B) THE PROPOSED METHOD

(A)

O group	P_I group	P_S group
$Cp1$-$Cp2$	$Bp1$-$Bp2$	$Pp8$-$Pp9$
$Bp6$-$Bp7$	$Pp1$-$Pp2$	$Dp1$-$Dp2$
$Ap6$-$Ap7$	$Pp4$-$Pp5$	$Cp4$-$Cp5$
	$Ap2$-$Ap3$	**$Fp2$-$Fp3$**
		$Tp1$-$Tp2$

(B)

O group	P_I group	P_S group
$Bp1$-$Bp2$	$Pp4$-$Pp5$	$Pp8$-$Pp9$
$Cp1$-$Cp2$	$Ap2$-$Ap3$	$Bp6$-$Bp7$
$Pp1$-$Pp2$	**$Dp1$-$Dp2$**	$Ap6$-$Ap7$
		$Cp4$-$Cp5$
		$Fp2$-$Fp3$
		$Tp1$-$Tp2$

From Table IV, we observe that both methods were able to correctly classify $Cp1$-$Cp2$ in the O group. Besides, they were also able to group properly all the P_S channels. Moreover, according to the expert, $Pp8$-$Pp9$ showed, contrary to the present seizure, a delayed involvement in other recorded seizures. This fact can explain here the classification of $Pp8$-$Pp9$ in the P_S group using both algorithms. Similarly, $Bp6$-$Bp7$ has shown a non-involvement behavior in other recorded seizures of the same patient. This may explain the fact that this channel was classified in the P_S group using the proposed algorithm. Now, the ERR-based method classified $Bp6$-$Bp7$ in the O group, which seems less appropriate. Following the expert's opinion, our method outperforms the original one in the classification of $Bp1$-$Bp2$, $Pp1$-$Pp2$ and $Dp1$-$Dp2$ channels. To conclude, the proposed approach appears attractive and more reliable in the identification of brain regions involved in the seizure onset, which is a crucial point from a therapeutic point of view.

IV. CONCLUSION

A robust ERR-based algorithm to infer causal interactions among brain regions in the context of epilepsy, through nonlinear system identification, was proposed in this paper. It mainly addressed the issue of spurious solutions often encountered in the original ERR-based method well-tailored to solve the nonlinear system identification problem. The proposed approach relies on the assumption of a sparse representation of the model coefficient vector. The well-known ADMM method combined with an optimal computation of the regularization parameter, at each iteration, was used to solve the optimization problem at hand. Results on both simulated and real iEEG data confirmed the reliability of the proposed approach over the conventional ERR-based one.

REFERENCES

[1] A. N. Almeida, V. Martinez, and W. Feindel, "The first case of invasive EEG monitoring for the surgical treatment of epilepsy: historical significance and context," *Epilepsia*, vol. 46, no. 7, pp. 1082–1085, 2005.

[2] W. Penfield and T. C. Erickson, *Epilepsy and Cerebral Localization*. Charles C. Thomas, 1941.

[3] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[4] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their applications to non-linear system identification," Tech. Rep. 2-30, Sheffield University, 1989.

[5] S. A. Billings, M. J. Korenberg, and S. Chen, "Identification of nonlinear output-affine systems using an orthogonal least squares algorithm," Tech. Rep. 1-11, Sheffield University, 1987.

[6] Y. Zhao, S. A. Billings, H. Wei, and P. G. Sarrigiannis, "Tracking time-varying causality and directionality of information flow using an error reduction ratio test with applications to electroencephalography data," *Physical Review E*, vol. 86, no. 5, pp. 1–11, 2012.

[7] P. G. Sarrigiannis *et al.*, "Quantitative EEG analysis using error reduction ratio-causality test; validation on simulated and real EEG data," *Clinical Neurophysiology*, vol. 125, pp. 32–46, 2014.

[8] Y. Zhao *et al.*, "Imaging of nonlinear and dynamic functional brain connectivity based on EEG recordings with the application on the diagnosis of Alzheimer's disease," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1571–1581, 2020.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[11] K. El Houari *et al.*, "Investigating transmembrane current source formulation for solving the ECG inverse problem," in *IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 371–375, 2018.

[12] N. Galatsanos and A. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.

[13] C. He, C. Hu, W. Zhang, and B. Shi, "A Fast adaptive parameter estimation for total variation image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 4954–4967, 2014.

[14] A. Fasoula, Y. Attal, and D. Schwartz, "Comparative performance evaluation of data-driven causality measures applied to brain networks," *Journal of Neuroscience Methods*, vol. 215, no. 2, pp. 1–20, 2013.

[15] N. Biggs, E. K. Lloyd, and R. J. Wilson, *Graph Theory 1736-1936*. Oxford University Press, 1986.